

εσταδιστιχ̄

εσταδιστιχ̄

Inferencia estadística / Estadística aplicada

UOC

ΕΣΤΑΔΙΣΤΙΧ

Apuntes



PEC 1: TEORÍA

1. LA DISTRIBUCIÓN NORMAL

ESTADISTIX

$$Z = \frac{x - \mu}{\sigma}$$

Una pizzería reparte pizzas a domicilio. Se ha estudiado que el tiempo de entrega es una variable aleatoria que sigue una distribución Normal con media $\mu = 41$ y desviación estándar $\sigma = 10$. Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es la probabilidad de que la pizza tarde más de 44,853 minutos?
- b) ¿Cuál es la probabilidad de que la pizza tarde entre 37,15 y 53,816 minutos?

Para resolver este ejercicio hay que usar la siguiente tabla con valores de la distribución normal estándar:

$P(Z < z)$	0,500	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950
Valor de z	0,000	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645

Nota: Introduce las respuestas como probabilidades en tanto por ciento (esto es, valores entre 0 y 100), sin decimales y sin el símbolo % al final.

Una empresa de distribución de productos congelados está estudiando la temperatura (en grados celsius) de un producto a su llegada al cliente final. Se sabe que la temperatura media de este producto sigue una Normal con media $\mu = -12,432$ y desviación estándar $\sigma = 12$. Con esta información, responde a las siguientes preguntas:

- **a)** ¿Cuál es la probabilidad de que llegue un producto con una temperatura de cero grados o superior?
- **b)** ¿Cuál es la temperatura que deja en la cola izquierda el 5% de la distribución?

Para resolver este ejercicio hay que usar la siguiente tabla con valores de la distribución normal estándar:

$P(Z < z)$	0,500	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950
Valor de z	0,000	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645

ESTADISTIX

Una clínica ha lanzado un estudio con un grupo de pacientes, con el objetivo de investigar el tiempo de recuperación, en días, tras un tipo de operación. El estudio ha revelado que la variable aleatoria X (tiempo de recuperación) sigue una distribución Normal con desviación estándar $\sigma = 7$. Con esta información, responde a las siguientes preguntas:

- **a)** Sabiendo que la probabilidad de que el tiempo de recuperación sea igual o superior a 43,894 días es de 0,2 (esto es, un 20%), ¿cuál es la media (μ) de la variable aleatoria X ?
- **b)** Otra clínica diferente quiere realizar el mismo estudio con sus pacientes, y lo que encuentran es que, en su caso, la media μ es la misma, pero la probabilidad de que el tiempo de recuperación sea igual o superior a 41,08 días es de 0,35 (esto es, un 35%). En el caso de esta segunda clínica, ¿cuál es la desviación estándar σ del tiempo de recuperación?

Para resolver este ejercicio hay que usar la siguiente tabla con valores de la distribución normal estándar:

$P(Z < z)$	0,500	0,550	0,600	0,650	0,700	0,750	0,800	0,850	0,900	0,950
Valor de z	0,000	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645

2. INTERVALO DE CONFIANZA

Explicación de clase sobre población y muestra:

MUESTRA	→	ESTADÍSTICOS:	\bar{x}	S_x	S_x^2	\hat{p}	n
POBLACIÓN	→	PARÁMETROS:	μ	σ_x	σ_x^2	p	N

Distribución muestral: es el modelo de probabilidad de un estadístico cuando conocemos su parámetro.

Error típico o estándar: es la desviación típica de la distribución muestral. Nos indica la dispersión que hay entre los distintos valores del estadístico que podrían salir al recoger una muestra de tamaño n.

$$EE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad EE_{\hat{p}} = \sqrt{\frac{p \cdot (1-p)}{n}} \quad C.A \text{ medias} \begin{cases} x \sim N \\ o \\ n \geq 30 \end{cases} \quad C.A. \text{ proporciones} \begin{cases} n \cdot p \geq 5 \\ n \cdot (1-p) \geq 5 \end{cases}$$

Intervalos de confianza, ME y valor crítico:

$$IC_{\text{medias con } \sigma \text{ conocida o } n \geq 30} \quad \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$IC_{\text{medias con } \sigma \text{ desconocida}} \quad \bar{x} - t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$IC_{\text{proporciones}} \quad \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} < \pi < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

$$IC_{\text{varianzas}} \quad \frac{(n-1) \cdot s^2}{\chi_{n-1; \alpha/2}^2} < \sigma^2 < \frac{(n-1) \cdot s^2}{\chi_{n-1; 1-\alpha/2}^2}$$

El departamento de educación de un país quiere saber cuál es la altura media de los estudiantes de un curso determinado. Para ello, selecciona una muestra aleatoria de $n = 70$ estudiantes, midiéndoles la altura (en centímetros). De la muestra se obtiene una media muestral $\bar{x} = 148$ y una desviación estándar muestral $s = 15$. Además, se constata que se desconoce la varianza poblacional, y para el análisis se considera una significación de $\alpha = 0,1$. Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es el análisis estadístico que hay que realizar en este caso?
- b) ¿Cuál es la distribución de probabilidad que hay que utilizar?
- c) ¿Cuál es el límite inferior del intervalo de confianza para el parámetro de interés?
- d) ¿Cuál es el límite superior del intervalo de confianza para el parámetro de interés?

Nota 1: Para calcular los valores críticos en las distribuciones de probabilidad, se recomienda la siguiente página web:

<https://www.geogebra.org/m/U5rw94DP>

Student glib 69

P(1.66724 ≤ X) = 0.05

En un país se celebra una jornada electoral, donde se vota en un referéndum con dos opciones: sí y no. Durante la jornada se realizan encuestas a pie de urna, en las que se pregunta por el voto a $n = 600$ votantes, obteniendo que 420 han votado que sí y 180 han votado que no.

El objetivo del estudio es obtener un intervalo de confianza del porcentaje de votantes que votan afirmativamente, considerando una significación de $\alpha = 0,01$. Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es el análisis estadístico que hay que realizar en este caso?
- b) ¿Cuál es la distribución de probabilidad que hay que utilizar?
- c) ¿Cuál es el límite inferior del intervalo de confianza para el parámetro de interés?
- d) ¿Cuál es el límite superior del intervalo de confianza para el parámetro de interés?

Nota 1: Para calcular los valores críticos en las distribuciones de probabilidad, se recomienda la siguiente página web:

<https://www.geogebra.org/m/U5rw94DP>

Normal μ 0 σ 1

P(2.57583 ≤ X) = 0.005

En una jornada electoral decisiva se presentan dos candidatos: el verde y el azul. Durante la jornada se realizan encuestas a pie de urna, en las que se pregunta por el voto a n votantes. Con una confianza del 95% se ha encontrado que el intervalo de confianza estimado para el porcentaje de votos del candidato verde es $[0,5693, 0,6299]$. Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es el análisis estadístico que hay que realizar en este caso?
- b) ¿Cuál es la distribución de probabilidad que hay que utilizar?
- c) ¿Cuál es el Margen de Error en la estimación?
- d) ¿Cuál es el tamaño de la muestra (n)?

Nota 1: Para calcular los valores críticos en las distribuciones de probabilidad, se recomienda la siguiente página web:

<https://www.geogebra.org/m/U5rw94Dp>

Normal μ 0 σ 1

$P(1.95996 \leq X) = 0.025$

ESTADISTIX



3. EJERCICIOS CON R

Una empresa quiere analizar la dispersión de la duración media, en días, de uno de sus productos. Específicamente, se quiere crear un intervalo de confianza para el parámetro de la dispersión (varianza). Para ello, la empresa selecciona una muestra con $n = 90$ productos, obteniendo una varianza muestral igual a $s^2 = 40$. Además, para realizar la inferencia estadística se considera una significación de $\alpha = 0,01$. Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es el análisis estadístico que hay que realizar en este caso?
- b) ¿Cuál es la distribución de probabilidad que hay que utilizar?
- c) ¿Cuál es el límite inferior del intervalo de confianza para el parámetro de interés?
- d) ¿Cuál es el límite superior del intervalo de confianza para el parámetro de interés?

Nota 1: Para calcular los valores críticos en las distribuciones de probabilidad, se recomienda la siguiente página web:

<https://www.geogebra.org/m/USrw94DP>

Importante: Para abrir el enlace, haz clic sobre el botón derecho del ratón y selecciona la opción de abrir enlace en una pestaña nueva.

Nota 2: Al hacer los cálculos, considera siempre un mínimo de tres decimales para los números decimales.

Nota 3: Introduce las respuestas redondeando a dos decimales y usando la coma (,) como separador decimal (no el punto).

El departamento de educación de un país quiere saber cuál es el peso medio de los estudiantes de un curso determinado. Para ello, selecciona una muestra aleatoria de n estudiantes, midiéndoles la altura (en centímetros). Esta muestra está recogida en la hoja de cálculo `Peso.xlsx`, que hay que descargar para realizar el ejercicio.

Además, se constata que se desconoce la varianza poblacional, y para el análisis se considera una significación de $\alpha = 0,1$.

Con esta información, responde a las siguientes preguntas:

- a) ¿Cuál es el valor crítico?
- b) ¿Cuál es Margen de Error?
- c) ¿Cuál es el límite inferior del intervalo de confianza para el parámetro de interés?
- d) ¿Cuál es el límite superior del intervalo de confianza para el parámetro de interés?

Nota 1: El siguiente enlace ofrece una descripción de las principales funciones estadísticas de Microsoft Excel, las cuales resultan de utilidad en la realización de este ejercicio.

<https://exceltotal.com/funciones/estadisticas/>

Importante: Para abrir el enlace, haz clic sobre el botón derecho del ratón y selecciona la opción de abrir enlace en una pestaña nueva.

Nota 2: Al hacer los cálculos, considera siempre un mínimo de tres decimales para los números decimales.

Nota 3: Introduce las respuestas redondeando a dos decimales y usando la coma (,) como separador decimal (no el punto).

Una empresa tecnológica quiere analizar la duración media, en minutos, de una batería para ordenadores portátiles. Específicamente, se quiere crear un intervalo de confianza para el parámetro de la varianza de esa duración. Para ello, la empresa selecciona una muestra con n baterías, y analiza su duración en minutos. Esta muestra está recogida en la hoja de cálculo *Duracion.xlsx*, que hay que descargar para realizar el ejercicio.

Además, para realizar la inferencia estadística se considera una significación de $\alpha = 0,05$.

Con esta información, responde a las siguientes preguntas:

- **a)** ¿Cuál es el valor crítico inferior?
- **b)** ¿Cuál es el valor crítico superior?
- **c)** ¿Cuál es el límite inferior del intervalo de confianza?
- **d)** ¿Cuál es el límite superior del intervalo de confianza?

Nota 1: El siguiente enlace ofrece una descripción de las principales funciones estadísticas de Microsoft Excel, las cuales resultan de utilidad en la realización de este ejercicio.

Un instituto de investigación en epidemiología realiza un ensayo clínico para evaluar un nuevo medicamento. Para eso, necesita estudiar qué parte de la población tiene un determinado tipo de anticuerpos (es decir, es seropositiva). Para ello, analiza una muestra de n pacientes, obteniendo un intervalo de confianza estimado para el porcentaje de casos positivos (es decir, con anticuerpos) igual a $[0,4114, 0,4886]$.

Con esta información, y usando los datos disponibles en *Ensayo.xlsx*, responde a las siguientes preguntas:

- **a)** ¿Cuál es Margen de Error?
- **b)** ¿Cuál es el valor crítico?
- **c)** ¿Cuál es el nivel de confianza, en tanto por ciento?
- **d)** ¿Cuál es el nivel de significación, en tanto por ciento?

Nota 1: El siguiente enlace ofrece una descripción de las principales funciones estadísticas de Microsoft Excel, las cuales resultan de utilidad en la realización de este ejercicio.

<https://exceltotal.com/funciones/estadisticas/>

Importante: Para abrir el enlace, haz clic sobre el botón derecho del ratón y selecciona la opción de abrir enlace en una pestaña nueva.

Nota 2: Al hacer los cálculos, considera siempre un mínimo de tres decimales para los números decimales.

Nota 3: En las preguntas a) y b), introduce las respuestas redondeando a dos decimales y usando la coma (,) como separador decimal (no el punto).

Nota 4: En las preguntas c) y d), introduce las respuestas en tanto por ciento y sin añadir el símbolo % al final.

PEC 1: RESOLUCIÓN

Una gran empresa dedicada a la comercialización online de productos cosméticos acaba de lanzar un artículo al mercado. Para poder obtener retroalimentación de los clientes, el departamento de marketing ha diseñado una encuesta y la ha distribuido entre los mismos. A partir de las diferentes preguntas de la encuesta se construyen dos variables:

- **Satisfacción con la entrega:** variable que refleja la satisfacción del cliente en relación al proceso de entrega del producto a domicilio, incluyendo la rapidez y el embalaje del producto.
- **Satisfacción con el producto:** variable cualitativa que muestra las estrellas que el cliente le otorga al producto, siendo una estrella la satisfacción más baja, y cinco estrellas la más alta.

Esta muestra está recogida en la hoja de cálculo [Encuesta.xlsx](#), que hay que descargar para realizar el ejercicio.

El estudio se compone de dos partes. En la primera de estas, se pretende construir un intervalo de confianza para la satisfacción media de los clientes con el proceso de entrega. Para esta primera parte del estudio se considera una significación de $\alpha = 0,05$, y además se constata, por estudios previos, que la varianza poblacional es $\sigma^2 = 196$.

Con esta información, responde a las siguientes preguntas:

- **a)** ¿Qué clase de inferencia tenemos que realizar?
- **b)** ¿Cuál es la media muestral de la satisfacción con la entrega?
- **c)** ¿Cuál es margen de error?
- **d)** ¿Cuál es el límite inferior del intervalo de confianza para el parámetro de interés?
- **e)** ¿Cuál es el límite superior del intervalo de confianza para el parámetro de interés?

La segunda parte del estudio se dedica a analizar la satisfacción del cliente con el producto.

Específicamente, se desea estudiar la proporción de clientes que otorgan al producto una satisfacción máxima (cinco estrellas) sobre el total. Para esta proporción se estima un intervalo de confianza igual a $[0,3047, 0,3953]$. Con esta información, responde a las siguientes preguntas:

- **f)** ¿Qué clase de inferencia hemos realizado?
- **g)** ¿Cuál es la proporción muestral de clientes que han otorgado cinco estrellas al producto, en tanto por uno?
- **h)** ¿Cuál es el valor crítico?
- **i)** ¿Cuál es el nivel de significación, en tanto por ciento?
- **j)** ¿Cuál es el nivel de confianza, en tanto por ciento?

Nota 1: al hacer los cálculos, considera siempre un mínimo de tres decimales para los números decimales.

Nota 2: introduce las respuestas redondeando a dos decimales y usando la coma (,) como separador decimal (no el punto).

Nota 3: en las preguntas i) y j), no introduzcas el símbolo del porcentaje (%) al final.

Este dossier está hecho para seguir la clase de prueba.

Más información en:

www.estadistix.com

**Y si tienes cualquier consulta,
escribenos un whatsapp al 644310902**

